Contents

1 Tools - Sampling Theory-univariate analysis

1.1 Tools for determining probabilities that a sample mean $\bar{x}$ or $\hat{p}$ is of a given value.

Assumptions: Population parameters are known.
Sample distribution is normal.

1. $\bar{x}$

   (a) For sampling distribution

   $$\mu_{\bar{x}} = \mu$$
   $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

   (b) Convert $\bar{x}$ to a z score using the following function

   $$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

   (c) Use the table in the back of the book to calculate the probability.

2. $\hat{p}$

   (a) Sample proportion is determined by:

   $$\hat{p} = \frac{x}{n}$$

   (b) For the sampling distribution

   $$\mu_{\hat{p}} = p$$
   $$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

   where q = 1 - p.
   If np > 5 and nq > 5, the sampling distribution can be approximated by the normal distribution.

   (c) Convert $\hat{p}$ into a z score and calculate the probability.

   $$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}}$$

   (d) Use the table in the back of the book to calculate the probability.

1.2 Margin of error

1. For point estimate $\bar{x}$

$$\pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

   If σ is unknown and n $\geq$ 30, one can substitute s for σ.

2. For point estimate $\hat{p}$

$$\pm 1.96\sqrt{\frac{pq}{n}}$$

   When p is unknown, margin of error can be estimated as

$$\pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

   Recall: $n\hat{p} > 5$ and $n\hat{q} > 5$

1.3 Interval estimator

1. General function - two tail test

$$\text{Point estimator} \pm z_{\frac{\alpha}{2}} * \text{Standard Error}$$

   (a) Population mean

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

   When n > 30

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

   (b) Population proportion

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

2. General function - left tail test (one-sided confidence interval)

$$\text{Point estimator} - z_{\alpha}\text{Standard Error}$$

3. General function - right tail test (one-sided confidence interval)

$$\text{Point estimator} + z_{\alpha}\text{Standard Error}$$

2 Values of $z_{\frac{\alpha}{2}}$ and z_α for given values of α.

Confidence	α	$\frac{\alpha}{2}$	$z_{\frac{\alpha}{2}}$	z_α
99.0%	0.010	0.0050	2.58	2.33
98.0%	0.020	0.0100	2.33	2.055
97.5%	0.025	0.0125	2.24	1.96
95.0%	0.050	0.0250	1.96	1.645
90.0%	0.100	0.0500	1.645	1.28

3 Tools - Sampling Theory-bivariate analysis

3.1 Properties of Sampling distribution of $(\bar{x}_1 - \bar{x}_2)$

1. Mean and Standard error

$$\mu_{(\bar{x}_1 - \bar{x}_2)} = \mu_1 - \mu_2$$

$$SE = \sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

2. Margin of Error

$$\pm 1.96\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

3. Confidence interval (two-tail)

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

If σ_1^2 and σ_2^2 are unknown, but both n_1 and n_2 are greater than or equal to 30, you can substitute the sample variances for the population variances.

3.2 Properties of Sampling distribution of $(\hat{p}_1 - \hat{p}_2)$

1. Mean and Standard Error

$$(\hat{p}_1 - \hat{p}_2) = \left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right)$$

$$\mu_{(\hat{p}_1 - \hat{p}_2)} = (p_1 - p_2)$$

$$SE = \sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

2. Margin of Error

$$\pm 1.96\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

3. Confidence Interval (two-tail)

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$n_1 p_1, n_1 q_1, n_2 p_2$ and $n_2 q_2 > 5$

4 Sample Size

Analysis	**Estimator**	**Minimum sample size**
Univariate	$\bar{x}$	$n \geq (z^2_{\alpha/2} \; * \; \sigma^2)/B^2$
	$\hat{p}$	$n \geq (z^2_{\alpha/2} \; * \; pq)/B^2$
Bivariate	$\bar{x}_1 - \bar{x}_2$	$n \geq (z^2_{\alpha/2} \; * \; (\sigma_1^2 + \sigma_2^2))/B^2$
	$\hat{p}_1 - \hat{p}_2$	$n \geq (z^2_{\alpha/2} \; * \; (p_1 q_1 + p_2 q_2)/B^2$

For the Bivariate functions $n_1 = n_2 = n$.
B is the acceptable margin of error.
If σ is not known, the sample standard deviation can be used or a value based on the range of the values divided by 4.

5 Decision tree - sample test for parameter

1. One Variable
   (a) Interval variable type $\bar{x}$
      i. n is greater or equal to 30 - **Sample Test for the mean(μ)-large sample**
      ii. n is less than 30 - - **Sample Test for the mean(μ)-small sample**
   (b) Proportion p, n is greater than 30 - **Sample Test for proportion(p)-large sample**
2. Two Variables
   (a) Interval variable type $\bar{x}$
      i. Equal variances ($\sigma_1^2 = \sigma_2^2$) or n $\geq$ 30
         A. n greater or equal to 30 - **Sample Test for the difference in means(μ_1-μ_2)-large sample**
         B. n less than 30 - - **Sample Test for the difference in means (μ_1-μ_2)-small sample**
         C. n less than 30, paired test - **Paired Differenced Test for the mean(μ_d)-small sample**
      ii. Unequal variances($\sigma_1^2 \neq \sigma_2^2$), n is less than 30 - **Sample Test for the difference in means (μ_1-μ_2)-small sample, unequal variances**
   (b) Proportion p, n is greater than 30 - **Sample Test for differences in proportion($p_1 - p_2$)-large sample**

6 Sample Test for the mean(μ)-large sample

Make Assumption
Interval variable, normal population, univariate analysis, large sample

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0 \text{ (two tailed test)}$$
$$H_a : \mu < \mu_0 \text{ or } \mu > \mu_0 \text{ (one tailed test)}$$

Obtain sampling distribution
Z distribution
μ_0 - known value given above

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \text{ or } \frac{s}{\sqrt{n}}$$

Choose rejection values from Table 14

Compute test statistics

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

Make a decision

7 Sample Test for the difference in means(μ_1-μ_2)-large sample

Make Assumption
Interval variables, normal population, bivariate analysis, large sample

$$H_0 : (\mu_1 - \mu_2) = D_0$$
$$H_a : (\mu_1 - \mu_2) \neq D_0 \text{ (two tailed test)}$$
$$H_a : (\mu_1 - \mu_2) < D_0 \text{ or } (\mu_1 - \mu_2) > D_0 \text{ (one tailed test)}$$

Obtain sampling distribution
Z distribution
D_0 - known values

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \text{ or}$$
$$\sigma_{\bar{x}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Choose rejection values from Table 14

Compute test statistics

$$z = \frac{(\bar{x_1} - \bar{x_2}) - D_0}{\sigma_{\bar{x}}}$$

Make a decision

8 Sample Test for proportion(p)-large sample

Make Assumption
Categorical variable, univariate analysis, large sample
Sampling satisfies the assumptions of a binomial experiment and n is large enough so that the sampling distribution of $\hat{p}$ can be approximated by a normal distribution

$$H_0 : p = p_0$$

$$H_a : p \neq p_0 \text{ (two tailed test)}$$

$$H_a : p < p_0 \text{ or } p > p_0 \text{ (one tailed test)}$$

Obtain sampling distribution
Z distribution
p_0 - known value

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0 q_o}{n}}$$

Choose rejection values from Table 14

Compute test statistics

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}$$

Make a decision

9 Sample Test for differences in proportion($p_1 - p_2$)-large sample

Make Assumption
Categorical variables, bivariate analysis, large samples. Sampling satisfies the assumptions of a binomial experiment and n is large enough so that the sampling distribution of $\hat{p}$ can be approximated by a normal distribution

$$H_0 : p_1 - p_2 = D_0$$

$$H_a : p_1 - p_2 \neq D_0 \text{ (two tailed test)}$$

$$H_a : p_1 - p_2 < D_0 \text{ or } p_1 - p_2 > D_0 \text{ (one tailed test)}$$

Obtain sampling distribution
Z distribution
D_0 - known values

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \text{ or}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$$

Choose rejection values from Table 14

Compute test statistics

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sigma_{\hat{p}}}$$

Make a decision

10 Sample Test for the mean(μ)-small sample

Make Assumption
Interval variable, normal population, univariate analysis, small sample

$$H_0 : \mu = \mu_0$$

$$H_a : \mu \neq \mu_0 \text{ (two tailed test)}$$

$$H_a : \mu < \mu_0 \text{ or } \mu > \mu_0 \text{ (one tailed test)}$$

Obtain sampling distribution
t distribution
μ_0 - known value given above

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Calculate degrees of freedom. Choose rejection values from Table 15
Degrees of freedom $\rightarrow$ (n-1)

Compute test statistics

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

Make a decision

11 Sample Test for the difference in means (μ_1-μ_2)-small sample, equal variances

Make Assumption
Interval variables, normal population, bivariate analysis, small sample, equal variances

$$\frac{\text{Larger } s^2}{\text{Small } s^2} \leq 3$$

$$H_0 : (\mu_1 - \mu_2) = D_0$$

$$H_a : (\mu_1 - \mu_2) \neq D_0 \text{ (two tailed test)}$$

$$H_a : (\mu_1 - \mu_2) < D_0 \text{ or } (\mu_1 - \mu_2) > D_0 \text{ (one tailed test)}$$

Obtain sampling distribution
t distribution
D_0 - known values

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Calculate degrees of freedom. Choose rejection values from Table 15
Degrees of freedom $\rightarrow (n_1 + n_2 - 2)$

Compute test statistics

$$t = \frac{(\bar{x_1} - \bar{x_2}) - D_0}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Make a decision

12 Sample Test for the difference in means (μ_1-μ_2)-small sample, unequal variances

Make Assumption
Interval variables, normal population, bivariate analysis, small sample, unequal variances

$$\frac{\text{Larger } s^2}{\text{Small } s^2} > 3$$

$$H_0 : (\mu_1 - \mu_2) = D_0$$

$$H_a : (\mu_1 - \mu_2) \neq D_0 \text{ (two tailed test)}$$

$$H_a : (\mu_1 - \mu_2) < D_0 \text{ or } (\mu_1 - \mu_2) > D_0 \text{ (one tailed test)}$$

Obtain sampling distribution
t distribution
D_0 - known values

Calculate degrees of freedom. Choose rejection values from Table 15
Degrees of freedom (rounded to the nearest integer)

$$df \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

Compute test statistics

$$t = \frac{(\bar{x_1} - \bar{x_2}) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Make a decision

13 Paired Differenced Test for the mean(μ_d)-small sample

Make Assumption
Interval variable, normal population, bivariate analysis, small sample

$$H_0 : \mu_d = \mu_0$$

where μ_d is average of the population differences.

$$H_a : \mu_d \neq \mu_0 \text{ (two tailed test)}$$

$$H_a : \mu_d < \mu_0 \text{ or } \mu_d > \mu_0 \text{ (one tailed test)}$$

Obtain sampling distribution
t distribution
μ_0 - known value given above (typically, it is zero)
$\bar{d}$ - average sample difference

Calculate degrees of freedom. Choose rejection values from Table 15
Degrees of freedom $\rightarrow$ (n-1)

Compute test statistics

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$$

Make a decision

14 Rejection values of z - one and two tailed regions

Two	Tailed	Region	One	Tailed	Region
p=0.10	p=0.05	p=0.01	p=0.10	p=0.05	p=0.01
1.645	1.96	2.58	1.28	1.645	2.33

15 Rejection values of t - one and two tailed regions

	Two	Tailed	Region	One	Tailed	Region
df	p=0.10	p=0.05	p=0.01	p=0.10	p=0.05	p=0.01
1	6.314	12.706	63.657	3.157	6.353	31.829
2	2.920	4.303	9.925	1.460	2.152	4.963
3	2.353	3.182	5.841	1.177	1.591	2.921
4	2.132	2.776	4.604	1.066	1.388	2.302
5	2.015	2.571	4.032	1.008	1.286	2.016
6	1.943	2.447	3.707	0.972	1.224	1.854
7	1.895	2.365	3.499	0.948	1.183	1.750
8	1.860	2.306	3.355	0.930	1.153	1.678
9	1.833	2.262	3.250	0.917	1.131	1.625
10	1.812	2.228	3.169	0.906	1.114	1.585
11	1.796	2.201	3.016	0.898	1.101	1.508
12	1.782	2.179	3.055	0.891	1.090	1.528
13	1.771	2.160	3.012	0.886	1.080	1.506
14	1.761	2.145	2.977	0.881	1.073	1.489
15	1.753	2.131	2.947	0.877	1.066	1.474
16	1.746	2.120	2.921	0.873	1.060	1.461
17	1.740	2.110	2.898	0.870	1.055	1.449
18	1.734	2.101	2.878	0.867	1.051	1.439
19	1.729	2.093	2.861	0.865	1.047	1.431
20	1.725	2.086	2.845	0.863	1.043	1.423
21	1.721	2.080	2.831	0.861	1.040	1.416
22	1.717	2.074	2.819	0.859	1.037	1.410
23	1.714	2.069	2.807	0.857	1.035	1.404
24	1.711	2.064	2.797	0.856	1.032	1.399
25	1.708	2.060	2.787	0.854	1.030	1.394
26	1.706	2.056	2.779	0.853	1.028	1.390
27	1.703	2.052	2.771	0.852	1.026	1.386
28	1.701	2.048	2.763	0.851	1.024	1.382
29	1.699	2.045	2.756	0.850	1.023	1.378

Interpretation of table. (Example, two tailed test, df=10)
If $|t| > 1.812 \rightarrow$ reject H_0 at p value (significance level) of 0.10.
If $|t| > 2.228 \rightarrow$ reject H_0 at p value (significance level) of 0.05.
If $|t| > 3.169 \rightarrow$ reject H_0 at p value (significance level) of 0.01.